The Interactive Effects of Problem Difficulty and Reinforcement of Social Behavior on Academic Performance in the Classroom

Wendy Lee Leys
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/masters_theses
Part of the Educational Psychology Commons

Recommended Citation
https://scholarworks.wmich.edu/masters_theses/2356

This Masters Thesis-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Master's Theses by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
THE INTERACTIVE EFFECTS OF PROBLEM DIFFICULTY
AND REINFORCEMENT OF SOCIAL BEHAVIOR ON ACADEMIC
PERFORMANCE IN THE CLASSROOM

by

Wendy Lee Leys

A Project Report
Submitted to the
Faculty of The Graduate College
in partial fulfillment
of the
Specialist in Education Degree

Western Michigan University
Kalamazoo, Michigan
August 1976
ACKNOWLEDGEMENTS

I would like to express my gratitude to the Portage Public Schools for allowing me the opportunity to work in their setting and especially to Mr. Bullard and Mr. Searl for their cooperation and suggestions. I would also like to thank Drs. Brian A. Iwata, Galen J. Alessi, and Howard E. Farris for the support and encouragement they have offered during my work on this project. Special appreciation goes to Dr. Iwata for his continued guidance and advice which proved invaluable throughout all stages of this study. I would also like to express my gratitude to Diane Nelles and Bob Weis who volunteered their time and energy as observers, fellow researchers, and helped me in countless other ways in the implementation of my project. Finally, a note of thanks must go to Carol Daisley who assisted me in the development of my academic materials on many occasions.

To all of these people, and many others, I gratefully acknowledge the assistance and guidance that made this project possible.

Wendy Lee Leys
INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)”. If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in “sectioning” the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from “photographs” if essential to the understanding of the dissertation. Silver prints of “photographs” may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Xerox University Microfilms
300 North Zeeb Road
Ann Arbor, Michigan 48106

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
LEYS, Wendy Lee
THE INTERACTIVE EFFECTS OF PROBLEM DIFFICULTY AND REINFORCEMENT OF SOCIAL BEHAVIOR ON ACADEMIC PERFORMANCE IN THE CLASSROOM.

Western Michigan University, Ed.S., 1976
Education, general

Xerox University Microfilms, Ann Arbor, Michigan 48106
TABLE OF CONTENTS

Chapter                                      Page

I    INTRODUCTION .........................................  1

II   EXPERIMENT I ............................................11
     Method ............................................... 11
     Subjects and Setting .............................. 11
     Observation ..................................... 11
     Arithmetic Materials and Measures .......... 13
     Reliability ..................................... 14
     Procedure ...................................... 15
     Experimental Design ............................. 17
     Results .......................................... 19

III  EXPERIMENT II ............................................32
     Method ........................................... 32
     Subjects and Setting ............................ 32
     Observation ................................... 32
     Arithmetic Materials and Measures .......... 32
     Reliability .................................... 33
     Experimental Design ............................ 35
     Results ......................................... 36

IV   DISCUSSION ..............................................48

V    REFERENCES ..............................................55
INTRODUCTION

The development and use of token reinforcement procedures has become a common occurrence in applied settings over the past ten years. A review of the research in this area indicates that token systems have been successfully employed with a variety of populations including psychiatric inpatients (Ayllon and Azrin, 1968; Atthowe and Krasner, 1968); mentally retarded adults and children (Girardeau and Spradlin, 1964; Birnbrauer and Lawler, 1974); classroom students (Staats, Finley, Minke, Wolf and Brooks, 1964; O'Leary and Drabman, 1971) and delinquents (Tyler and Brown, 1968). These studies have repeatedly demonstrated the effectiveness of token reinforcement in working with differing populations and for producing changes in a wide variety of behaviors (Kazdin and Bootzin, 1972; O'Leary and Drabman, 1971).

Within the classroom setting, token procedures are of particular importance in maximizing teacher effectiveness and student performance. In a review of token reinforcement techniques, Walker and Buckley (1974) outline the advantages of token systems in classroom management:

1. The distribution of reinforcers in a manner that is easy for the teacher and not disruptive to the ongoing class activities.
2. The provision of a variety of backup reinforcers to meet the needs of the individual student.
3. The use of systematic changes in the schedule of reinforcement required by the specific stage of learning.
4. The strengthening of conditioned reinforcers more
typically found in the classroom (e.g., praise, grades, etc.).

Traditional academic incentives such as praise, grades, and advancement may not be reinforcers for all students, thus explaining the little effect these consequences have on the behavior of some students. A token economy provides an "artificial" situation in which a variety of activities and materials that will serve as reinforcers may be made available contingent upon appropriate behavior. Through frequent pairings with these reinforcers, the more traditional incentives can be made to serve as generalized conditioned reinforcers which can be used effectively by teachers in motivating and rewarding student behavior in the classroom.

Studies have examined the effects of token systems in modifying a variety of classroom behaviors such as disruption (Bailey, Wolf, and Phillips, 1970; Barrish, Saunders, and Wolf, 1969; Iwata and Bailey, 1974; Kaufman and O'Leary, 1972; O'Leary, Becker, Evans, and Saudargas, 1969), study behavior (Surratt, Ulrich, and Hawkins, 1969), and academic performance (Ayllon, Layman, and Burke, 1972; Ayllon and Roberts, 1974; Dalton, Rubino, and Hislop, 1973; Lahey and Drabman, 1974). In these studies, one class of behavior was selected for intervention and token reinforcement was delivered contingent upon behavior appropriate to that category. Results of these and other studies illustrate that a variety of classroom behaviors can be modified by contingent token reinforcement.

Upon closer examination of these studies, an unanswered question becomes apparent. Since token procedures have been found effective
in modifying behavior, a decision must be made regarding those behaviors to be selected for intervention. Two classes of behavior are considered of prime importance in the classroom: social behavior that is compatible with accomplishing academic tasks, and academic performance that meets the instructional objectives of the teacher. A frequent complaint voiced by teachers requesting professional assistance in their classroom concerns high degrees of disruptive behavior. It is often the case that behavior problems within the classroom are severe enough to necessitate the initial selection of disruptive behavior as the first priority for change. As the level of appropriate social behavior increases and the rate of disruptions declines, more attention may be directed toward the increasing of academic performance which is the terminal goal in any classroom program.

One possible direction to be followed when evaluating the selection of the behavior to be consequated in a token system is that of investigating the generalization of treatment effects to other, non-targeted, behaviors. Kazdin (1973) suggests that the future development of reinforcement procedures must be concerned with not only the direct modification of targeted behavior but also with the concurrent effects of such procedures on other behaviors within the individual's repertoire. This attention to non-target behavioral measures will encourage the continued investigation and development of behavior management strategies designed to most effectively increase inappropriate and adaptive behavior of more than one specified
A number of studies have provided information on the spread of effects in a token system from the target category to other behaviors but the data in this area are not consistent. In an investigation of the effects of reinforcement of academic behavior on disruptive behavior, two studies may serve to illustrate these contradictory findings. Dalton et al. (1973) awarded retarded students with a token for each correct response to orally presented questions from the Distar Arithmetic and Language Kits. The token group was compared to a control group receiving no tokens for academic performance and conduct. A significant increase in correct responses was reported for the token group but no difference in conduct between the groups was observed. The results of this study, showing no carry-over of academic contingencies to social behavior, are contradicted by a second study investigating the same question. Ayllon and Roberts (1974) presented data taken during the implementation of a token system in which points exchangeable for backup reinforcers were delivered for academic accuracy on a reading workbook task. When the point system was implemented, the percent of correct responses increased and disruptive behavior decreased from the baseline levels. These data were presented as a mean of the five subjects observed in the study and are not completely representative due to the fact that one of the five subjects did not show significant academic improvement and persisted in high rates of disruption for a longer duration than the other subjects. Neither of the above studies
presented data on the rate of problem completion and it is therefore difficult to determine the total impact of the increase in percentage correct (i.e., student may have improved in accuracy scores by doing fewer problems), but it does appear that academic performance can be improved by direct intervention. The generalization of this effect to non-target behavior, however, is unclear.

Studies in which disruptive behavior has been selected as the target behavior have produced similarly inconsistent data. The dependent variables identified by these studies as measures of academic performance range from subjective analysis to objective measures of the number of problems completed and percent correct. Surratt et al. (1969) reported a token program in which four 1st grade subjects received reinforcers for the amount of time spent working. Under reinforcement conditions, the amount of time on task increased and the teacher reported that the subjects completed more problems with a higher degree of accuracy. No objective data, however, were presented to substantiate this suggested improvement. In another study, O'Leary et al. (1969) implemented a token system in a 2nd grade class with points delivered for following the rules outlining appropriate social behavior. The token system was effective in significantly reducing the rate of rule violations and produced a mean gain of 1.5 years, as measured by the California Achievement Test. The effect of the token procedures on this gain cannot be clearly determined, but the authors suggested that such an above average gain would most likely be due to experimental manipulations.
of point delivery for appropriate behavior. Kaufman and O'Leary
(1972) measured academic improvement in a more controlled manner.
The social and academic behavior of subjects in two classrooms were
compared following the onset of a reward procedure in one class and
a response cost procedure in the other. In both classes, tokens
were contingent on following the class rules. Disruptive behavior
was reduced for both groups and no significant differences were
found between the reward and cost procedure. Completion of SRA
Power Builders was used as one measure of academic performance and
the results obtained from these data indicated a significant, but
not consistent, improvement for subjects in both classes. In the
reward class, the number of power builders increased for four sub-
jects, remained the same for one, and decreased for three subjects
when the token system was in effect. Seven subjects in the cost
class increased in the number completed during the token phase, while
one subject decreased. The second measure of academic performance
taken in this study was test scores on the Wide Range Achievement
Test. For both classes, a mean gain of .6 years was measured from
the onset to the completion of the study (3.5 months). This above
average gain was again attributed to token reinforcement of social
behavior. A second study, also investigating the effects of rein-
fforcement versus cost token procedures (Iwata and Bailey, 1974),
replicated Kaufman and O'Leary's findings that the number of problems
completed (math in this study) increased when tokens were delivered
for appropriate behavior. Accuracy data were not included in the
former study, but Iwata and Bailey reported that, although tokens contingent on social behavior increased the rate of academic performance, accuracy remained relatively unchanged across all experimental conditions. Bailey et al. (1970) reported similar results in a study investigating the use of a token reinforcement procedure on social and academic performance. The first part of this study involved five delinquent boys who received points for appropriate social conduct, including study behavior, which were exchangeable for daily home privileges. When these backup reinforcers were made available, the frequency of study behavior was increased and the number of rule violations was decreased. Academic measures were taken for the number of math workbook problems completed and the percent correct. Again, as in the Iwata and Bailey (1974) study, the number of problems completed increased under token conditions, but in this case the number of problems correct not only failed to increase, but declined steadily throughout the study.

The question of the spread of effects (generalization) to non-target behaviors has also been investigated by means of direct comparison. Ferritor, Buckholdt, Hamblin, and Smith (1972) reported a study in which points were given first for academic performance, next for attending behavior, and finally for both academic and attending behavior. The academic task selected for this study was completion of math worksheets. Results indicated that conduct contingencies did not have a significantly positive effect on academic performance nor did academic contingencies maintain appropriate

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
classroom behavior. It was found that the simultaneous presentation of both contingencies was necessary to control the academic and social behaviors. However, Sulzer, Hunt, Ashby, Konairski, and Krams (1971) presented data that do not support this conclusion. In their study, points given for items correct were compared to points delivered for on-task behavior in relation to their effects on both academic and social performance in reading and spelling tasks. Results indicated that during the reading sessions both types of point contingencies positively affected the non-target behavior. Spelling tasks, however, produced different results. When points were delivered for items correct, both the percent of accuracy and the percent of on-task behavior increased. This carry-over was not found when points were delivered for on-task behavior; during this condition the accuracy declined to that of the initial baseline. The authors concluded that the generalization of effects was less probable in token systems designed to consequate appropriate social behavior and that reinforcement of items correct was a more effective method for producing changes in both academic performance and conduct.

In light of the above, it appears that whether it is more preferrable to select academic or social behavior as the target variable is not a clear "either-or" situation. Although improved academic performance is the explicit goal of all token procedures in the classroom, two problems exist which limit the practical use of selecting academic behavior as the initial concern. The first problem, mentioned previously, occurs when classroom behavior problems have
become so severe that they cannot be even temporarily ignored. Second, tokens contingent upon academic performance require that each student has the prerequisite skills necessary to meet the criteria established for reinforcement. Extensive amounts of time and effort on the part of the teacher may be necessary to determine this level of prerequisite skill, which may even necessitate the development of individualized curricula. Thus, in situations where inappropriate social behavior is the first priority for change, it would seem reasonable to target appropriate behavior for reinforcement and to increase the probability of concurrent positive change in the non-target behavior of academic performance if at all possible.

The purpose of this study is to investigate selected variables within the classroom setting that may affect generalization of treatment effects from a target to a non-target behavior. The variable chosen for examination is the level of difficulty of the material being worked on by the student during the token session. Two studies (Bailey et al., 1970; Iwata and Bailey, 1974) have suggested that increases in difficulty level over the course of their studies may have been a factor which limited increases in accuracy when tokens were delivered for social behavior. This variable is not often mentioned in the literature and is potentially important in that the difficulty of material, regardless of prerequisite skills, may affect the number of problems completed as well as the accuracy with which they are completed. If such is the case, a teacher can provide academic work that tends to maximize academic success, thus
augmenting the effect of token reinforcement for social behaviors compatible with academic achievement. In this manner, the shift of emphasis from social to academic behavior has already begun - the student is being reinforced for behaviors compatible with academic behavior while he is completing a greater percent of work accurately. The definition of difficulty selected to be investigated by this study is that of the response effort required to successfully complete academic work and the probability of error determined by differing presentations of assigned problems. The possible interactive effects of different problem format and reinforcement for social behavior will be studied in Experiment I and Experiment II, and the questions to be specifically investigated by these two experiments are:

1. What is the effect on on-task behavior and the frequency of rule violations when points are delivered for following class rules?

2. What is the effect on academic performance when a token system is introduced for appropriate social behavior?

3. Is the difficulty of the assigned material (response effort) one variable determining the extent of generalization to academic performance when tokens are delivered for appropriate social behavior?
EXPERIMENT I

Method

Subjects and Setting

Ten 5th grade students attending a public elementary school served as subjects. They were selected on the basis of the teacher reports indicating high rates of disruptive behavior and low rates of assignment completion. Six students of this group were identified for observation of social behavior while measures of academic performance were taken for all ten subjects. At the onset of the study the group consisted of six boys and four girls; following Session 10 one boy moved to another school district and was eliminated from the study.

The study was conducted in an unused classroom in the school. At the beginning of each morning Math period, the subjects left their regular classroom and went with the experimenter to this room for the session.

Observation

Daily observations of social behavior of the six target subjects were taken during the first ten minutes of the math session. A 10-sec observe - 5-sec record interval recording procedure was used. One of two observers was present each day and recorded from the front of the classroom. An observer recorded the behavior of three target
subjects for one minute and then moved to the remaining three subjects, alternating in this manner throughout the 10-min session. Observed verbal interaction of the teacher with any of the three target subjects or to the group was also recorded at this time.

A tape recorder with ear phones was used to mark the beginning and end of the recording intervals. Observers wore sunglasses at all times while they were recording to make it difficult for the subjects to determine who they were watching.

The two categories of student behavior observed in this study were:

1. Rule Violations: Noncompliance with any of the class rules, including:
   a. Out-of-Seat: Movement of the child from his chair when not permitted and/or requested by the teacher. Movement was defined as the child's buttocks not in contact with the seat of the chair.
   b. Talking without Permission of Instructor: Any unpermitted vocalization that was audible to the observer.
   c. Engaging in Motor Behavior that Disturbs Other Students: Hitting desk, physical contact with other students or other student's desk, and other noise-making activities that attract the attention of other students.

2. On-Task: Attention to materials or teacher and/or hand raised for the entire 10-sec interval.

These two categories were not mutually exclusive and it was possible for both to be scored during any 10-sec interval. If neither category of behavior was observed during an interval, a third category, "other", was scored.

Teacher verbal behavior was recorded during each session to ensure that social reinforcement and feedback remained constant throughout the study. The two categories of teacher behavior observed
were:

1. Teacher Approval: Any verbal response of approval to either the individual student being observed or to the group that was audible to the observer.

2. Teacher Disapproval: Any verbal disapproval to the individual student being observed or to the group that was audible to the observer.

Arithmetic Materials and Measures

A method of column addition was taught to the subjects and used as the measure of academic performance. This method has been described as a "low fatigue algorithm" on the basis of the operations it requires (Hutchings, 1972). It involves the addition of only two digits at a time of no greater value than nine, in contrast to the traditional algorithm which requires the cumulative addition of two or more digits. The algorithm can be described as follows:

Half-space notation uses numerals of no more than a half-space in height to record the sum of two digits. With half-space notation, the units portion of the sum of two digits is written at the lower right of the bottom digit and the tens portion is written at the lower left of the bottom digit...The ones portion of the column sum is always the same as the ones portion of the last two-digit sum...The tens portion of the column sum is always the same as the number of tens recorded at the left of the column. These are simply counted [Example A]...For a column in some multicolumn exercise...the total number of tens, is no longer written in the tens place of the first column's sum but instead at the top of the next column at the left [Example B] (Hutchings, 1976, p. 221).
Problems were randomly generated on a computer and presented as ditto worksheets. Two types of problem arrays were employed. Low Response effort problems were defined as 6 by 2 arrays, \((T = 11\) binary operations), while High Response effort problems were defined as 6 by 6 arrays, \((T = 35\) binary operations).

Measures of performance on these problems were calculated daily. The number of columns completed during the 10-min session and the percentage correct were recorded.

**Reliability**

Frequent reliability checks were made throughout the study to assess the accuracy of observer recordings. Separate reliabilities were calculated for each category of social behavior by dividing the total number of agreements by the total number of agreements.
plus disagreements and multiplying by 100. An agreement was
scored if both observers recorded the same behavior within the same
10-sec interval. A disagreement was scored if one observer recorded
the behavior and the other did not.

Reliability data on social behavior were obtained for 20 of
the 36 experimental sessions, yielding an overall mean of 94% and
a range of 77% - 100%. The following means were obtained for each
category observed: on-task - 94%; rule violations - 93%; teacher
approval - 100%; and teacher disapproval - 100%.

Reliability on academic measures was checked at least once
during each phase by an independent observer who scored the papers
and recorded the number of problems completed and percent correct
on a separate score sheet making no marks on the original paper.
Following this, the experimenter scored the papers and a comparison
was made of the two sets of scores. Reliability was calculated by
dividing the lower score by the higher score for number of problems
completed and the percent correct and multiplying by 100. Reliability
checks on academic measures were made seven times during the study
and 100% agreement was found for both the number of columns com-
pleted and the percentage correct.

Procedure

Prebaseline

Data were taken for 30 minutes each day for five days prior
to the onset of the study on the rate of On-Task and Rule Violations
exhibited by ten students in the class. The group of ten children served as the subjects of this study and their academic performance was measured for the duration of the experiment. The six subjects of this group who were selected on the basis of Prebaseline data as the most highly disruptive were also targeted for observation of social behavior.

**Baseline**

On the day before the study began the teacher told the students that they would be divided into two groups during the math period. The students were told that the selection was made on a random basis and that the division of the class was done to insure more individualized instruction. On the first day of Baseline, the ten subjects selected for the experimental group went to a second classroom where the math period was conducted for the duration of the study.

A rule board was posted in the front of the classroom and the class rules were explained at this time. The rules were:

1. Be in your seat at 11:30 (the beginning of the session);
2. Stay in your seat;
3. Raise your hand for help; and
4. Work by yourself.

These rules were read at the beginning of each math session of the study.

Following a review of the rules, each subject was given a worksheet with column addition problems on it. When all papers had been handed out, the subjects were told to begin work. After one sheet was completed, the subject was given another sheet. At the end of
ten minutes, the subjects were told to stop work and the papers were collected.

Points

On Session 14 the subjects were told that points would be awarded for following the class rules and that these points would be exchangable for free time on Fridays. At the end of every 5-min period, the experimenter recorded the number of points earned by each child on a point chart posted at the front of the room. Two points could be earned during each 5-min interval and were awarded as follows: 2 points were earned if the subject followed all the rules for the entire 5-min; 1 point was awarded if a rule was broken once; and 0 points were awarded if more than one rule violation occurred. When the points were recorded on the chart, the reason for any loss was briefly explained. A student could earn a total of four points each day and the points earned were recorded on a cumulative point chart until Friday (total possible 16 points). On Friday each point was exchangable for one minute of free time. If a student was absent on one day, the points earned on the following day were doubled so as to not penalize the student's possible attainment of free time. The following were selected by the subjects as free time activities: 1. Quizmo (math bingo) with candy and cookie prizes; 2. Going to the library; and 3. Going to the gym or outside.

Experimental Design

A reversal design (Baer, Wolf, and Risley, 1968) was used to
evaluate the effects of reinforcement on the rate of On Task-behavior and Rule Violations. Baseline I was employed to establish the current level of academic and social behavior. The points' phase was implemented to evaluate the effects of reinforcement for appropriate social behavior on both the rate of On-Task and Rule Violations. Baseline conditions were reinstated for the final eight days of the study in order to establish functional control of the dependent variables.

A multielement design (Ullman and Sulzer-Azaroff, 1975) was employed to assess the interactive effects of reinforcement of social behavior on academic performance. The days on which the subjects worked on either Low Response (6 x 2 column problems) or High Response problems (6 x 6 column problems) were alternated on a random basis with no more than two consecutive days in any one condition. This design made it possible to assess the differential effect on performance on the two formats of academic materials (6 vs. 2 column displays) when reinforcement for appropriate social behavior was delivered.
Results

Social Behavior

Figure 1 presents mean percentages of intervals in which On-Task and Rule Violations were exhibited by the six target subjects throughout the experimental conditions of this study. During Baseline, the means for On-Task behavior and Rule Violations were 63% and 40%, respectively. The data during this phase were highly variable and the daily group means ranged from 30% to 82% for On-Task and from 15% to 60% for Rule Violations. When points for rule following were instituted, the percentage of intervals of On-Task behavior increased to 100% on the first day and remained high throughout the phase, yielding a group mean of 97%. There was a concurrent decrease in Rule Violations during the points condition, dropping to a mean of 0%. Data during the points condition appeared stable showing little variability with the exception of the last On-Task data point which was lower than any time during this phase. Upon withdrawal of the point system, On-Task behavior decreased to a mean of 78% and Rule Violations increased to 28%. As in the first Baseline, the data were highly variable during this phase. No systematic differences in the two categories of social behavior were found to be related to the type of academic materials that were presented. As can be seen in Table 1, individual subject data were quite consistent with the group means.
Figure 1: Mean percent of intervals of On-Task behavior and Rule Violations across experimental conditions during Experiment I. On-Task measures are indicated by closed circles and Rule Violations by open circles. Both measures are labeled by arrows.
MEAN PERCENT OF INTERVALS FOR ON-TASK AND RULE VIOLATIONS

BASELINE

POINTS

BASELINE

MEAN PERCENT OF INTERVALS FOR ON-TASK AND RULE VIOLATIONS

SESSIONS

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 1: Mean percent of intervals of On-Task and Rule Violations for individual subjects across experimental conditions during Experiment I.
TABLE 1

MEAN PERCENT OF INTERVALS OF ON-TASK AND RULE VIOLATIONS FOR INDIVIDUAL SUBJECTS ACROSS EXPERIMENTAL CONDITIONS DURING EXPERIMENT I

<table>
<thead>
<tr>
<th>Subject</th>
<th>Baseline</th>
<th>Treatment</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Mean</td>
<td>On-Task</td>
<td>63</td>
<td>97</td>
</tr>
<tr>
<td>Rule Violations</td>
<td>40</td>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td>1</td>
<td>On-Task</td>
<td>80</td>
<td>99</td>
</tr>
<tr>
<td>Rule Violations</td>
<td>38</td>
<td>0</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>On-Task</td>
<td>85</td>
<td>100</td>
</tr>
<tr>
<td>Rule Violations</td>
<td>12</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>On-Task</td>
<td>54</td>
<td>97</td>
</tr>
<tr>
<td>Rule Violations</td>
<td>41</td>
<td>1</td>
<td>37</td>
</tr>
<tr>
<td>4</td>
<td>On-Task</td>
<td>58</td>
<td>96</td>
</tr>
<tr>
<td>Rule Violations</td>
<td>39</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>5</td>
<td>On-Task</td>
<td>65</td>
<td>98</td>
</tr>
<tr>
<td>Rule Violations</td>
<td>60</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>On-Task</td>
<td>28</td>
<td>91</td>
</tr>
<tr>
<td>Rule Violations</td>
<td>54</td>
<td>0</td>
<td>52</td>
</tr>
</tbody>
</table>
**Arithmetic Performance**

Low Response (LR) and High Response (HR) conditions were randomly presented for a total of 18 days in each condition. Figure 2 presents the mean number of columns completed and percentage correct on both LR and HR days. During Baseline I, there was a gradual upward trend for the number of columns completed during both LR and HR conditions. There was little difference between the number of columns completed during LR and HR conditions yielding group means of 27 and 28, respectively. A 7% difference in percentage correct was found with a group mean of 84% on LR days and 77% on HR days. When points were instituted for social behavior, the number of columns completed increased to a mean of 48 on LR days and to 44 on HR days. The mean percentage of problems correct rose from 84% to 89% on LR days and from 77% to 87% on HR days. During this phase, the data stabilized and each session mean varied little from the group mean. A return to Baseline conditions resulted in a greater variability in the data but little actual change in the academic means. The mean number of columns completed decreased by 2 (mean = 46) on LR days and by 5 (mean = 39) on HR days. The mean percentage of columns correct decreased by only 1% on LR days dropping from 89% to 88% and 4% on HR days, from 87% to 83%.

Individual subject data presenting the number of columns completed and the percentage correct on LR days are shown in Table 2. These data were inconsistent with the group means for several subjects. The number of columns completed by Subject 3 during the LR
Figure 2: Mean number of columns completed and percent correct on High Response and Low Response days for Experiment I. High Response days are indicated by open circles and Low Response days by closed circles. Both measures are labeled by arrows.
Table 2: Mean percent of accuracy and columns completed for individual subjects across experimental conditions on LR days during Experiment I.
### TABLE 2

MEAN PERCENT OF ACCURACY AND COLUMNS COMPLETED FOR INDIVIDUAL SUBJECTS ACROSS EXPERIMENTAL CONDITIONS ON LR DAYS DURING EXPERIMENT I

<table>
<thead>
<tr>
<th>Subject</th>
<th>Baseline</th>
<th>Treatment</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percent Correct</td>
<td>91</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>Number Completed</td>
<td>40</td>
<td>61</td>
</tr>
<tr>
<td>1</td>
<td>Percent Correct</td>
<td>81</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>Number Completed</td>
<td>38</td>
<td>54</td>
</tr>
<tr>
<td>2</td>
<td>Percent Correct</td>
<td>88</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>Number Completed</td>
<td>20</td>
<td>45</td>
</tr>
<tr>
<td>3</td>
<td>Percent Correct</td>
<td>94</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>Number Completed</td>
<td>26</td>
<td>39</td>
</tr>
<tr>
<td>4</td>
<td>Percent Correct</td>
<td>89</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>Number Completed</td>
<td>27</td>
<td>52</td>
</tr>
<tr>
<td>5</td>
<td>Percent Correct</td>
<td>78</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>Number Completed</td>
<td>11</td>
<td>31</td>
</tr>
<tr>
<td>6</td>
<td>Percent Correct</td>
<td>92</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>Number Completed</td>
<td>44</td>
<td>74</td>
</tr>
<tr>
<td>7</td>
<td>Percent Correct</td>
<td>84</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>Number Completed</td>
<td>28</td>
<td>42</td>
</tr>
<tr>
<td>8</td>
<td>Percent Correct</td>
<td>85</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>Number Completed</td>
<td>27</td>
<td>36</td>
</tr>
</tbody>
</table>
condition increased by more than half when points for social behavior were instituted with the percentage of accuracy increasing by 4%. When points were removed, the number of columns completed decreased to a mean of 36 for this subject and percentage of accuracy dropped from 92% to 72%. There were similar increases in the number of columns completed per session for Subject 6 and Subject 7 during the points condition and, for both subjects, this increase was not maintained when points were withdrawn.

Individual subject academic data on HR days are presented in Table 3. Again, inconsistent effects across subjects can be seen. The number of columns completed and the percent accuracy increased on HR days at the beginning of the points condition for Subject 3 and Subject 6. This increase did not maintain when points were removed and both the number of columns completed and percent correct decreased when points for appropriate social behavior were removed. The number of columns completed also increased for Subject 7 when points were in effect and decreased when they were removed but the percent accuracy for this subject gradually decreased throughout the study regardless of experimental manipulations on social behaviors.

**Teacher Behavior**

No instances of either teacher approval or disapproval were observed during any condition of this study. The session was conducted during an independent work period and interactions between students and the teacher were not observed as either reinforcing or punishing.
Table 3: Mean percent of accuracy and columns completed for individual subjects across experimental conditions on HR days during Experiment I.
<table>
<thead>
<tr>
<th>Subject</th>
<th>Baseline</th>
<th>Treatment</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Percent Correct</td>
<td>87</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>Number Completed</td>
<td>37</td>
<td>62</td>
</tr>
<tr>
<td>2</td>
<td>Percent Correct</td>
<td>64</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>Number Completed</td>
<td>33</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>Percent Correct</td>
<td>28</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>Number Completed</td>
<td>6</td>
<td>38</td>
</tr>
<tr>
<td>4</td>
<td>Percent Correct</td>
<td>96</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>Number Completed</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td>5</td>
<td>Percent Correct</td>
<td>77</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>Number Completed</td>
<td>25</td>
<td>46</td>
</tr>
<tr>
<td>6</td>
<td>Percent Correct</td>
<td>68</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>Number Completed</td>
<td>7</td>
<td>20</td>
</tr>
<tr>
<td>7</td>
<td>Percent Correct</td>
<td>92</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>Number Completed</td>
<td>42</td>
<td>68</td>
</tr>
<tr>
<td>8</td>
<td>Percent Correct</td>
<td>71</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>Number Completed</td>
<td>21</td>
<td>32</td>
</tr>
<tr>
<td>9</td>
<td>Percent Correct</td>
<td>84</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>Number Completed</td>
<td>22</td>
<td>36</td>
</tr>
</tbody>
</table>
EXPERIMENT II

Method

Subjects and Setting

The ten subjects of Experiment I served as the subjects of this study with the same six students targeted for observation of social behavior. This second experiment was conducted for a 20-min period each day in the same classroom as in Experiment I, and immediately followed the daily sessions for Experiment I.

Observation

All methods of observation were identical to those used in Experiment I with the exception that data were taken for 20-min each day following the 10-min of Experiment I.

Arithmetic Materials and Measures

Problems from the Heath Elementary Mathematics text (Heath & Company) were used as the measure of academic performance in this study. The book provides pre- and post-tests enabling the selection of assignments based on the individual performance levels of the student. Worksheets were written for each assignment from the text using identical problems and presenting them in a multiple choice or matching format. These worksheets were used as the Low Response problems while problems selected directly from the text (with no
additional prompts) were defined as High Response problems. Both story and basic computational problems were included in each format. At the end of each 20-min session, the papers were placed by each student in his individual folder and were graded by the experimenter following the session. The number of problems completed during the session and the percent correct were recorded daily for each student.

Reliability

Reliability was checked frequently and calculated for each category of behavior in the same manner as in Experiment I. Reliability checks on social behavior (conduct) were taken during 22 of the 39 experimental sessions of this study yielding an overall mean of 92% and a range from 78% to 100%. Agreement percentages for separate categories were: On-Task - 91%; Rule Violations - 90%; Teacher Approval - 100%; Teacher Disapproval - 100%. Reliability data were obtained on the academic measures of problems completed and percentage correct once during each experimental phase yielding a mean of 100% for problems completed and 99% for percentage correct.

Procedure

Prebaseline

A pre-test for the Math unit to be begun on the first day of Baseline was administered to all students in the class prior to the onset of Baseline. On the basis of these test results, individual assignment sheets for the chapter were written. Other conditions
during this phase were identical to those in Experiment I.

**Baseline**

On the first day of Baseline, the subjects were given an individual math folder following the 10-min session of Experiment I. The experimenter explained that the next work would be problems from the text and that on certain days this work would be done on worksheets and on other days it would be directly from the text. The daily assignment was clipped to the front of the folder and marked to indicate where the subject was to begin work. When finished with one assignment, the subject raised a hand. The experimenter then checked to see that it was completed and gave the next scheduled assignment. At the end of the 20-min session, subjects were instructed to stop work and put the papers in the folders. The work was graded by the experimenter following the session and returned in the folder the next day.

**Points**

The point system and reinforcers were presented as described in Experiment I. During each session of Experiment II, the subject could earn a total of eight possible points. Daily points were recorded on a cumulative point chart displaying the points awarded for both experiments. If a student earned the maximum number of points for all sessions during the week (16 - Exp. I + 32 - Exp. II) he was given a special treat (candy or cookie) at the beginning of
the Friday session and then allowed to choose a reinforcing activity. Subjects who failed to earn the total number of points were able to engage in free time activities for only the number of minutes earned.

The point system was implemented on the same day for Experiments I and II. In order to strengthen the reinforcing value of points, the point system was initially introduced on a Friday and the subjects were allowed to exchange their earned points (12 possible) for free time on Monday following the work session. For the remainder of this phase, the points earned on Monday through Thursday were exchanged for free time on Friday.

**Experimental Design**

A reversal and multielement design were again used to assess the effects of reinforcement of appropriate social behavior and the possible interactive effect on academic performance. The reversal design provided a means of evaluating the effects of reinforcement for social behavior on the rate of appropriate social behavior. Low Response effort (worksheets) and High Response effort problems (work directly from the text) were alternated in the same pattern as in Experiment I. The use of a multielement design permitted the investigation of how these two levels of materials were differentially affected by the increase in rate of appropriate social behavior.
Results

Social Behavior

The group mean percentages of intervals in which On-Task and Rule Violations were observed for the six target subjects are presented in Figure 3 and, as Table 4 shows, results for individuals were consistent with this group mean data. During Baseline I, the mean for On-Task was 49% and 52% for Rule Violations. When the Points phase was introduced, behavior increased immediately to an overall mean of 91%. The percentage of Rule Violations decreased at the onset of this condition to a mean of one percent. Upon the removal of points (Baseline II), On-Task behavior decreased by 30% to a mean of 60% and Rule Violations increased to a mean of 46% which was slightly lower than that of Baseline I.

Arithmetic Performance

The mean numbers of problems completed and percentages correct for LR and HR days are presented in Figure 4. Throughout all conditions, the number of problems completed per session and the percentage correct were generally higher on LR days. During the Baseline phase, a mean of 48 problems was completed per session on LR days and a mean of 18 problems on HR days. The percentage of accuracy during this phase was 81% on LR days and 63% on HR days. When points were introduced contingent upon appropriate social behavior, the number of LR problems completed and percentage correct decreased

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 3: Mean percent of intervals of On-Task behavior and Rule Violations across experimental conditions during Experiment II. On-Task measures are indicated by closed circles and Rule Violations by open circles. Both measures are labeled by arrows.
Table 4: Mean percent of intervals of On-Task and Rule Violations for individual subjects across experimental conditions during Experiment II.
# TABLE 4

**MEAN PERCENT OF INTERVALS OF ON-TASK AND RULE VIOLATIONS FOR INDIVIDUAL SUBJECTS ACROSS EXPERIMENTAL CONDITIONS DURING EXPERIMENT II**

<table>
<thead>
<tr>
<th>Subject</th>
<th>Baseline On-Task</th>
<th>Treatment On-Task</th>
<th>Baseline Rule Violations</th>
<th>Treatment Rule Violations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>61</td>
<td>95</td>
<td>84</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>74</td>
<td>98</td>
<td>85</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>33</td>
<td>98</td>
<td>85</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>45</td>
<td>87</td>
<td>40</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>56</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>50</td>
<td>95</td>
<td>79</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>62</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>88</td>
<td>30</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>65</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4: Mean number of problems completed and percent correct on High Response and Low Response days for Experiment II. High Response days are indicated by open circles and Low Response days by closed circles. Both measures are labeled by arrows.
slightly, yielding group means for this phase of 42 problems completed and 79% correct. Throughout this phase, there was a general descending trend for the number of LR problems completed. The mean number of problems completed per session on Hr days increased to 27 and mean accuracy increased by 1% to 64% when points were delivered for rule compliance. When points were withdrawn (Baseline II), the number of LR problems completed and the percentage correct continued to decrease with a mean of 27 completed per session and 67% accuracy for this phase. The mean number of HR problems completed during Baseline II decreased to 21 and the mean percentage of accuracy for HR problems completed during this phase dropped to 52%. For both LR and HR conditions, the data measured on academic performance remained variable throughout the Points and Baseline II phases.

As can be seen in Tables 5 and 6, individual subject data for academic performance on LR and HR days was consistent with the presented group means with several exceptions. Subject 3 showed an increase in the number of problems completed on LR days and a return to Baseline rate when points for appropriate social behavior were withdrawn. Subjects 5, 6, and 7 showed increases in both the mean number of HR problems completed per session and the mean percentage of accuracy when points were delivered for social behavior. For all three of these subjects, performance on HR days declined to Baseline rates or lower, when points were withdrawn.

Teacher Behavior

As in Experiment I, no instances of teacher approval or disapproval were observed during any phases of the study.
Table 5: Mean percent of accuracy and problems completed for individual subjects across experimental conditions on LR days during Experiment II.
TABLE 5

MEAN PERCENT OF ACCURACY AND PROBLEMS COMPLETED FOR INDIVIDUAL SUBJECTS ACROSS EXPERIMENTAL CONDITIONS ON LR DAYS DURING EXPERIMENT II

<table>
<thead>
<tr>
<th>Subject</th>
<th>Baseline Percent Correct</th>
<th>Baseline Number Completed</th>
<th>Treatment Percent Correct</th>
<th>Treatment Number Completed</th>
<th>Baseline Percent Correct</th>
<th>Baseline Number Completed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>80</td>
<td>53</td>
<td>69</td>
<td>55</td>
<td>67</td>
<td>39</td>
</tr>
<tr>
<td>2</td>
<td>76</td>
<td>53</td>
<td>64</td>
<td>44</td>
<td>50</td>
<td>39</td>
</tr>
<tr>
<td>3</td>
<td>74</td>
<td>28</td>
<td>73</td>
<td>54</td>
<td>37</td>
<td>21</td>
</tr>
<tr>
<td>4</td>
<td>94</td>
<td>33</td>
<td>90</td>
<td>14</td>
<td>78</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>53</td>
<td>86</td>
<td>69</td>
<td>67</td>
<td>66</td>
<td>30</td>
</tr>
<tr>
<td>6</td>
<td>97</td>
<td>14</td>
<td>89</td>
<td>9</td>
<td>65</td>
<td>12</td>
</tr>
<tr>
<td>7</td>
<td>80</td>
<td>68</td>
<td>61</td>
<td>73</td>
<td>61</td>
<td>60</td>
</tr>
<tr>
<td>8</td>
<td>92</td>
<td>43</td>
<td>96</td>
<td>44</td>
<td>68</td>
<td>17</td>
</tr>
<tr>
<td>9</td>
<td>92</td>
<td>47</td>
<td>94</td>
<td>26</td>
<td>89</td>
<td>14</td>
</tr>
</tbody>
</table>

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 6: Mean percent of accuracy and problems completed for individual subjects across experimental conditions on HR days during Experiment II.
TABLE 6
MEAN PERCENT OF ACCURACY AND PROBLEMS COMPLETED FOR INDIVIDUAL SUBJECTS ACROSS EXPERIMENTAL CONDITIONS ON HR DAYS DURING EXPERIMENT II

<table>
<thead>
<tr>
<th>Subject</th>
<th>Baseline Percent Correct</th>
<th>Treatment Percent Correct</th>
<th>Baseline Percent Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>71</td>
<td>51</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>38</td>
<td>43</td>
</tr>
<tr>
<td>2</td>
<td>63</td>
<td>49</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>25</td>
<td>38</td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>45</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>94</td>
<td>77</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>16</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>56</td>
<td>70</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>41</td>
<td>23</td>
</tr>
<tr>
<td>6</td>
<td>60</td>
<td>71</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>62</td>
<td>70</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>39</td>
<td>20</td>
</tr>
<tr>
<td>8</td>
<td>87</td>
<td>72</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>41</td>
<td>23</td>
</tr>
<tr>
<td>9</td>
<td>75</td>
<td>74</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>21</td>
<td>11</td>
</tr>
</tbody>
</table>
DISCUSSION

Research has repeatedly demonstrated that token systems may effectively change the rates of the behaviors for which tokens are delivered. The question remains, however, as to what degree the reinforcement of one category of behavior will generalize to other, non-reinforced behaviors. The fact that generalization has been achieved in certain studies, but not in others, suggests that variables other than mere reinforcement of a target response might be phenomena of considerable importance. One such variable in academic settings might be the nature of the academic material being worked on while a token system for conduct behavior is in effect. The present study was designed to investigate one of the many possible variables affecting the rate of generalization to non-selected behavior. The variable chosen was that of response effort (high or low) necessary for correct completion of arithmetic problems. Point delivery contingent upon the students following class rules produced the expected results of increased on-task behavior and a sharp reduction in the frequency of rule violations. This marked improvement in conduct behavior, however, demonstrated little control over the mean academic performance of the subjects in either HR or LR conditions. In Experiment I, a general increasing trend was noted for both the number of columns completed and the percent accuracy over time. Experiment II produced no real changes across baseline and point conditions. The material format did exert some
control over the subjects' academic performance independent of the
delivery of points. This control is only slightly demonstrated in
Experiment I, but it can be clearly seen that in Experiment II the
subjects completed more problems with a higher percent accuracy under
LR conditions.

It is essential to consider Experiments I and II independently
when examining the possible factors which limited the generalization
of effect of conduct contingencies to academic behavior. In Experi­
ment I, there were no real differences between HR and LR work other
than the number of columns presented by each problem. The selection
of columns completed as the dependent variable was done in order to
facilitate the comparison of HR and LR data using an equivalent mea­
sure. Cumulative error (using an incorrect bridging number resulting
in an incorrect column sum) was greater on HR days and, thus, the
percentage correct was the only measure which was differentially
affected by problem format. Two possible factors may have been re­
sponsible for the failure of this academic measure to show change
under point conditions on either HR or LR days. The general increasing
trend during baseline for columns completed and percentage correct
on both HR and LR days suggests the existence of non-programmed rein­
forcers affecting the rate of academic behavior. The novelty of
Hutching's method, the change to a new room for the math session,
and inter-student competition for getting the most problems completed
may have all served as reinforcers that were not controlled by this
study. It is also possible, considering the high rate of column
completion and accuracy during baseline, that these measures reached a ceiling level and could not have increased when the points condition was implemented. The overall high rate of behavior and only slight increase in columns completed during the points condition and slight decrease during the second baseline may reflect these two factors of non-programmed reinforcers and possible ceiling effects.

A ceiling effect may have also contributed to the failure to show differential effects on academic performance on LR and HR days in Experiment II. The purpose of this second study was to present materials using a different definition of easy (LR) and difficult (HR) format than that employed in Experiment I. The high rate of problem completion and percent accuracy on LR days during the initial baseline suggests that the student performance during the LR condition may have reached maximum levels during baseline and, again, could be only minimally affected by the delivery of points for appropriate conduct.

The variability in the academic data and the failure to demonstrate improvement in performance in Experiment II can also be considered in light of the academic materials used in both LR and HR conditions. Bailey et al. (1970) and Iwata and Bailey (1974) have suggested that one possible explanation for the limited increase in accuracy exhibited by their subjects was the lack of prerequisite skills necessary for improvement. Many math skills are cumulative and the two chapters covered during this experiment relied on basic computational skills in addition, subtraction, multiplication, and
division. Many questions asked of the teacher during work sessions were related to these basic prerequisites, suggesting deficient skills in those areas necessary for appropriate academic performance. Throughout this study, daily assignments differed in the tasks required of the student (word versus basic computational problems, addition versus multiplication problems, etc.) and, perhaps, the variability of data across time in Experiment II may reflect differing prerequisites required for task completion. Thus, the present results may have partially reflected a problem of skill acquisition as opposed to skill maintenance. If prerequisite skills are lacking, an incentive system of any type may be only minimally effective in the improvement of academic performance relying on these basic skills.

This study was not able to provide conclusive evidence in support of a causal variable that may determine the extent to which token reinforcement affects non-target behaviors. However, it does provide several suggestions for further research in this area. The choice and control of materials at clearly different levels of difficulty merits further consideration in the investigation of the interactive effects these levels of academic materials may have with token reinforcement of conduct behavior. From the results of these two experiments, it is clear that a methodological problem to be solved is how to choose material which will not be too easy so as to create a ceiling effect during baseline, nor so difficult that students do not have the prerequisite skills necessary to increase
their performance. This would necessitate the delivery of a series of pre-tests measuring student performance at different levels of the selected materials in order to establish a stable rate of pre-baseline performance at a variety of these levels. The designation of two levels of these materials as 'easy' and 'difficult' may then be accomplished with the assurance that neither level is endangered by possible ceiling or cellar effects. Materials with normed and consistent levels of difficulty such as the SRA Power Builders, may be used to simplify this determination, but the procedure of individually assessing the pre-baseline rates of student performance will be essential to use these materials effectively.

This study investigated only one possible variable affecting the rate of generalization of effect of token systems on non-targeted behavior. Many others remain uninvestigated. The selection of the behavior to be reinforced presents another variable that may be important. Both on-task behavior and following the class rules have been frequently selected as the categories of behavior to be consecrated in a token system. The choice of one or the other of these behaviors may have an important effect on the extent to which generalization to academic behavior occurs. The reinforcement of the absence of disruptions, as done in this study, places no direct contingency on the rate of on-task behavior exhibited by the student. In effect, the concurrent increase in on-task behavior when rule violations decreased shown in both experiments of this study, was, in itself, a side effect of the points system. The subjects did not have to be
on task to receive reinforcement, they simply had to cease from violating the class rules. The differing Baseline rates of on-task behavior and rule violations across subjects and the fact that those subjects exhibiting the highest rates of rule violations demonstrated the greatest amount of generalization to academic performance, suggests another question. Is it the case that a contingency for on-task behavior will be more powerful than a contingency for the absence of rule violations in affecting behavior for subjects who do not initially exhibit extremely high rates of disruptive behavior? For the student who is engaged in a high rate of disruptive behavior, which is incompatible with remaining on task, it may be that a decrease in the frequency of this behavior is bound to produce change in academic behavior to some degree. However, for the student who rarely engages in such disruptive behavior, a contingency established for following the rules is less likely to change his behavior and, thus, less likely to change the level of his academic performance. Directly reinforcing the behavior of on-task may prove to be a more efficient target for a token system if generalization to academic behavior is desired.

A second variable which has not been examined in token research investigating the spread of effect of conduct behavior when a token or point system is in effect for academic performance is that of the academic materials used and the criteria for reinforcement. Reading tasks require a more sustained period of attention to complete successfully than do individual math problems. Thus, the choice of
subject material may affect the generalization to conduct behavior as a result of the requirements of the academic task itself. A second variable that may be of considerable importance in the generalization of academic contingencies to the rate of appropriate conduct behavior is the amount of work required of the student for reinforcement. A series of long (rather than short) assignments will require the subject to remain on task for a longer duration and may indirectly affect the rate of inappropriate behavior that may occur without the loss of reinforcers. The level of criteria established for reinforcement of academic performance (e.g., 100% versus 60%) is perhaps another factor which will affect the rate of appropriate social behavior. These variables may be instrumental in the degree to which the generalization of points delivered for academic performance to appropriate social behavior occurs, and, therefore, merit further investigation.

In summary, the question of whether contingencies established for one category of behavior will generalize to a second category of behavior does not appear to be one which will be answered easily. Characteristics of token systems themselves must be investigated and the important variables within them, that increase the potential for this generalization, identified. It will be only then that the development of token systems maximizing the effect of contingent token delivery will be able to be developed with the broadest range of effect possible.
REFERENCES


Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.